



TITLE:

ニューロ \mathbf{DP} による生産
ラインの最適制御 (不確実性の下で
の意思決定の数理)

AUTHOR(S):

大野, 勝久

CITATION:

大野, 勝久. ニューロ \mathbf{DP} による生産ラインの最適制御 (不確
実性の下での意思決定の数理). 数理解析研究所講究録 2003, 1306: 73-
82

ISSUE DATE:

2003-02

URL:

<http://hdl.handle.net/2433/42808>

RIGHT:

ニューロ DP による生産ラインの最適制御

名古屋工業大学 大野勝久(Katsuhisa Ohno)
Nagoya Institute of Technology

1. はじめに

日本 OR 学会創立 40 周年記念事業「統合オペレーション」特別研究プロジェクトの一環として、中部支部を中心に特設研究グループ「ジャストインタイム (JIT) 生産システム」(特設 G2 と略称されている)が、約 50 名のメンバーで活動を続けてきた。特設 G2 の趣旨は、「来るべき近未来グローバル生産・物流システムとして、あくまでもわが国固有の JIT 生産システムを基本とし、IT 技術を有機的に取り入れたグローバル生産・物流システムを構築し、その優位性を明らかにしたい。すなわち、本研究グループでは、IT 技術で武装した JIT 生産・物流システムが、近未来グローバル生産・物流システムとして最有力であることを示したい。」ことであった。

JIT(just in time)生産システムにおける最も革新的な考え方が、「後工程引き取り、後補充生産方式(いつ、何を、どれだけ必要かが最も早く、正確にわかる後工程が、使った分だけを前工程に引き取りに行き、前工程は引き取られた分だけを生産し、補充する)」である。プル(pull)方式とも呼ばれ、かんばんはこの方式における情報伝達・制御手段である。本発表の目的は、このかんばん方式を最適制御の観点から見直し、JIT 生産システムの進化を模索することである[1]。

まず次章で、[2]を参考に生産管理方式に関する研究を概観し、JIT との比較を紹介する。ついで、著者らが行った生産ラインの最適発注・生産政策を求める研究を述べ、人工知能の分野において強化学習(Reinforcement Learning)[3]とも呼ばれている、ニューロ・ダイナミックプログラミング(Neuro-Dynamic Programming, NDP) [4, 5]を用いて得られる準最適制御政策と最適化されたかんばん方式との数値比較を示す。

2. 生産管理方式と制御政策

2.1 生産管理・制御方式

1) 受注生産方式 (make-to-order system)

注文を受けてから必要な生産活動に入る生産ラインであり、製番管理方式、部品中心生産システム、生産座席システム等、納期と生産期間の関係で種々の管理方式が行われている。しかし最も単純なものは、注文を受ける毎に、必要な原材料は調達されたものとして先頭工程へ生産指示をだすものであり、通常の直列待ち行列あるいは待ち行列ネットワークとして定式化される。

2) 基点在庫方式 (base stock system)

この方式は、在庫管理における Clark and Scarf[6]によるエシェロン在庫 (echelon stock) の考え方にに基づき、在庫管理では[7]等で古くから知られている。ここでエシェロン在庫とは、その工程を含めて下流全てに存在する在庫量の和である。基点在庫方式は、あらかじめ各工程の基準在庫量を定めておき、その工程のエシェロン在庫位置 (エシェロン在庫 + 発注残) がその基準在庫量を下まわれば、基準在庫量まで生産する方式である。したがって、初期状態における各工程のエシェロン在庫位置を基準在庫量に設定しておけば、需要により最終製品が引きとられる毎に、全工程へ同時に生産指示が出される。

3) MRP(material requirements planning)

MRP(資材所要量計画)は、1960 年代から米国で開発されてきた生産管理方式である。対象となる品目を独立需要品目と従属需要品目に区分し、生産活動の全てをタイム・バケット(time bucket)と呼ばれる時間区間に対して計画し、そのタイム・バケット内に行われるように管理する。各タイム・バケットで生産すべき独立需要品目の生産量を与える基準生産計画と部品表に基づいて、必要となる部品量を計算し、各品目の使用可能在庫量からその発注・生産指示を

与える。

4) かんばん方式 (kanban system)

詳細は[8]等を参照されたい。

5) OPT(optimized production technology)

OPTは、イスラエルの物理学者 Goldratt 博士によって 1970 年代後半から開発されてきた生産スケジューリングソフトである。OPT は、固有の評価指標であるスループット(=売上げ一資材費)の向上とリードタイムの短縮、最適在庫水準の維持等の目標を達成するための最善策を、ボトルネック工程に着目して計画するソフトである。OPT では具体的な最適化手法は公表されていなかったが、その発展形である TOC(theory of constraints: 制約条件の理論)で明らかにされたボトルネック工程を最大限に活用するスケジューリング手法が、DBR(drum, buffer, rope)である。ここで、ドラムはボトルネック工程の生産に全工程が同期すること、バッファはボトルネック工程が仕掛品不足で止まらないように在庫を持つこと、ロープは先頭工程が進みすぎないようにボトルネック工程の生産に同期して原材料を投入することを意味している。

6) CONWIP[9]

かんばん方式の代替として提案されたプル方式であり、生産ライン内における総 WIP (work in process, 仕掛品) を常に一定 (CONstant) に保持する方式である。従って、最終製品が顧客に引き取られる毎に先頭工程に生産指示が出され、後は工程順に加工される。

[2]においては、これら方式がすべて PAC システム ([9]等参照) の例として定式化され、PAC システムの近似的な性能評価法が与えられている。しかし、各方式の比較あるいは最適制御政策については述べられていない。以下 1) から順に関連した研究を紹介する。

1) については[10]に、待ち行列ネットワークにおける先着順 (FIFO) サービスや多品種単一工程における $c\mu$ 則の最適性等が紹介されている。ここで $c\mu$ 則とは、品種 j の平均加工時間が $1/\mu_j$ 、単位時間当たりの遅れ費用が c_j で与えられたとき、 $c_j\mu_j$ が最大の品種を加工する政策である。また、Ohno and Ichiki[11]は、単一品種多工程直列生産ラインにおいて、注文の到着がポアソン過程に従い、各工程の加工時間が指数分布に従う (M/M 型と呼ぶ) ときの生産率の最適制御問題をマルコフ決定過程として定式化し、修正政策反復法を提案している。数値結果として、2 工程における最適制御政策

を与え、3 工程に対する計算時間等を示している。

2) については、その最適性が離散時間有限期間問題にたいして[6, 12]に示されている。ただし、最適性はその評価関数、制約条件等に依存して変化し、その詳細は直接論文を参照されたい。また、Decroix and Arreola-Risa[13]は、多品種無限期間問題にたいして基点在庫方式の最適性を示し、Rubio and Wein[14]はその最適基準在庫量がみたすべき条件を導き、Jackson 待ち行列ネットワークにたいしてその決定法を示している。さらに、Glasserman and Tayur[15]は、基点在庫方式のもとでの生産ラインの安定条件を導き、Chen[16]はエシェロン在庫を知るための情報の価値について論じている。

6) にたいして Spearman and Zazanis[17]は、単一品種生産ラインに対して「プル方式が MRP より優れており、かんばん方式のスループットは CONWIP のスループットを越えない」ことを示している。さらに、Muckstadt and Tayur[18, 19]は、「かんばん方式は同じスループットを CONWIP よりも少ない平均在庫量で達成し、一方 CONWIP はかんばん方式よりも変動の少ないスループットを、より少ない最大在庫量で与える」ことを示している。Yang[20]は、これらの結果が多品種生産ラインに対して成り立つかどうかを明らかにするために、6 品種 5 工程 M/M 型生産ラインのシミュレーションを実行し、CONWIP、1 枚かんばん(引き取りかんばん)、2 枚かんばんの各種性能(平均客待ち時間、平均在庫量、平均トリップ数)を生産方式、かんばん枚数、引き取り周期、品種選択則、移動政策を因子とする分散分析を行っている。結論は、「CONWIP は、かんばん方式よりも小さな平均客待ち時間、平均在庫量、平均トリップ数を達成するが、かんばん方式よりも広い在庫スペースを要求すること」である。

2.2 最適制御政策

Akella and Kumar[21]は、機械故障 (down と呼ぶ) と修理を考慮し、それ以外は確定的な単一品種単一機械工程において、在庫費用と品切れ費用からなる総費用を最小化する生産率の最適制御問題を論じている。生産可能なとき (up と呼ぶ) の最適制御政策は、最適在庫レベルを下まわれば最大生産率で生産し、最適在庫レベルに達すれば需要率と同じレベルに下げ、上まわれば停止する政策である。さらに Beilecki and Kumar[22]は、同じ単一品種単一機

械工程において、不確定な生産環境にもかかわらず、製品在庫をもたない政策が最適となる状況が存在することを示している。

Ryzin and Gershwin[23]は、各工程が up-down の 2 状態を独立なマルコフ連鎖としてとる、[21]と同様な単一品種 2 工程直列生産ラインにおいて、総費用を最小化する生産率の最適制御問題を論じ、系統的な数値結果の分析から 2 次元状態空間における最適制御政策なパターンを示している。Veatch and Wein[24]は、単一品種 2 工程 M/M 型生産ラインにおける生産率 μ の最適制御政策を論じ、ゼロ在庫政策、ゼロ製品在庫政策、工程 2 の非遊休政策が各々最適となる条件を導き、基点在庫政策が最適になりえないことを示している。また数値例により、基点在庫政策、かんぱん方式、固定バッファの順で最適政策に近いことも示している。

3. 生産ラインの最適制御

第 1 工程が外注工場等から部品を購入し、単一品種の製品を完成させる M 工程生産ライン(図 1)を考える。工程 i , $i=1, \dots, M$, の発注, 納入は各期首に行われ, 前工程は一定の納入リードタイム L_i 期後に受注した部品を納入する。工程 i の部品の最大在庫量を $I_{\max,i}$, 製品の倉庫容量を $J_{\max,i}$, 公称の生産能力を c_i とおく。しかし, チョコ停等のため c_i は達成できず, n , $n=1, 2, \dots$, 期における生産能力 $c_i(n)$ は, 各期独立に同一の分布に従うものとし, その最小値を $c_{i,\min}$ とする。また, 最終製品にたいする n 期の需要量 $D(n)$ は, 互いに独立で同一の分布に従うものとし, その最小値と最大値を D_{\min} , D_{\max} とし平均を D とおく。満たされなかった需要は次期に繰り越されるものとする。[25, 26]では, 故障の発生は無視できるものとし, 単一工程の最適発注・生産政策を論じている。

第 i 工程は, 第 n 期首において部品在庫量 $I_i(n)$ と製品在庫量 $J_i(n)$ を持つものとし, それら

システム全体の情報に基づいて, その期の部品発注量 $O_i(n)$, 製品生産量 $P_i(n)$ を決定する。 $J_i(n)$ の負の値は工程 $(i+1)$ の発注の繰り越し (品切れ) を意味している。そして, n 期首における工程 $i-1$ から i への納入量を $Q_i(n)$ とおく。

この生産ラインにたいして, 単位期間あたりの平均総費用を最小化する最適発注・生産政策を求める問題を考える。費用としては, 部品および製品の在庫費用および品切れ費用を考慮することとする。すなわち,

C_i^I : 各期における工程 i の部品在庫費用/個
 C_i^J : 各期における工程 i の製品在庫費用/個
 C_i^B : 各期における工程 i の繰り越し費用/個
 B_i : 各期における工程 i の繰り越し発生費用/回である。

第 n 期首における生産ラインの状態 s_n は, 各工程 i における第 $(n-L_i+1)$ 期から第 $(n-1)$ 期までの発注量および, 部品在庫量と製品在庫量のベクトルによって表される。すなわち,

$$s_n = (O_1(n-L_1+1), \dots, O_1(n-1), \dots, O_1(n-L_1+1), \dots, O_M(n-L_M+1), \dots, O_M(n-1), I_1(n), J_1(n), \dots, I_M(n), J_M(n)) \quad (1)$$

である。ここで $L_i=1$ の工程 i にたいしては, その期の発注量が次期に納入されるため, 発注量 O_i の情報は不要である。したがって, 全ての工程 i で $L_i=1$ ならば

$$s_n = (I_1(n), J_1(n), \dots, I_M(n), J_M(n)) \quad (2)$$

である。これら可能なすべての状態 s_n からなる状態空間を S とおく。

状態 $s_n \in S$ における工程 i の可能な発注量 $O_i(n)$ と生産量 $P_i(n)$ の集合は, 最大在庫量と生産能力の制限から各々次式で与えられる。

$$K_i^O(s_n) = \{0, \dots, I_{\max,i} - I_i(n) - \sum_{l=1}^{L_i-1} O_i(n-l)\}, \quad i=1, \dots, M \quad (3)$$

$$K_i^P(s_n) = \{0, \dots, \min\{I_i(n), C_i, J_{\max,i} - J_i(n)\}\}, \quad i=1, \dots, M-1. \quad (4)$$

最終工程 M にたいしては, その後工程は市場であり, 可能な生産量の集合は, 最終製品の倉

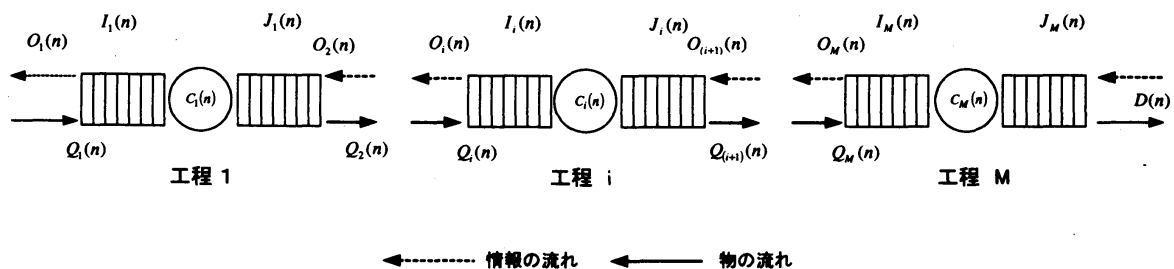


図 1 M 工程生産ライン

庫容量と需要の最小値を用いて次式で与えられる。

$$K_M^P(s_n) = \{0, \dots, \min\{I_M(n), C_M, J_{\max M} - J_M(n) + D_{\min}\}\} \quad (5)$$

すなわち、状態 s_n でとりうる決定

$a = (O_1(n), P_1(n), \dots, O_i(n), P_i(n), \dots, O_M(n), P_M(n))$ は $O_i(n) \in K_i^O(s_n)$, $P_i(n) \in K_i^P(s_n)$, $i=1, \dots, M$ を満たさなければならない。そして、(3)~(5)式で与えられる各工程の可能な発注量と生産量の集合の直積を $K(s_n)$ で表すことにすれば、 $a \in K(s_n)$ であり、政策 f は、各状態 s における可能な決定 $f(s)$ の集合 $\{f(s) \in K(s); s \in S\}$ である。

政策が決定されれば、次の期首の状態は以下のように定められる。

$$I_i(n+1) = I_i(n) + Q_i(n) - P_i'(n) \quad i=1, \dots, M \quad (6)$$

$$J_i(n+1) = J_i(n) + P_i'(n) - O_{i+1}(n - L_{i+1} + 1), \quad i=1, \dots, M-1 \quad (7)$$

$$J_M(n+1) = J_M(n) + P_M'(n) - D(n), \quad (8)$$

ここで $P_i'(n)$ は n 期の実際生産量であり、

$$P_i'(n) = \min\{P_i(n), C_i(n)\} \quad (9)$$

で与えられる。また、各工程における納入量は次式で表わされる。

$$Q_i(n) = \min\{O_i(n - L_i) + [-J_{i-1}(n-1)]^+, P_{i-1}'(n-1) + [J_{i-1}(n-1)]^+\}, \quad i=1, \dots, M, \quad (10)$$

ここで、 $[x]^+ = \max(0, x)$, $J_0(n) = 0$, $P_0'(n) = I_{\max,1}$ である。そして、状態 s_n で決定 a をとったとき、次期に状態 s_{n+1} へ推移する確率は、生産能力および需要分布を用いて以下のように与えられる。

$$p(s_n, s_{n+1}, a) = \begin{cases} \Pr\{C_i(n) = c_i, i=1, \dots, n, D(n) = d, s_{n+1} = (Q(n - L_1 + 2), \dots, Q(n), \dots, O_M(n - L_M + 2), \dots, O_M(n), I_1(n) + Q_1(n) - \min\{P_1(n), c_1\}, J_1(n) + \min\{P_1(n), c_1\} - O_2(n - L_2 + 1), \dots, I_M(n) + Q_M(n) - \min\{P_M(n), c_M\}, J_M(n) + \min\{P_M(n), c_M\} - d) \text{ のとき} \\ (C_{i,\min} \leq c_i \leq C_i, i=1, \dots, M, D_{\min} \leq d \leq D_{\max}) \\ 0, \quad \text{上記以外} \end{cases} \quad (11)$$

さらに、状態 s_n で決定 a をとったときの n 期における直接費用は、

$$r(s_n, a) = \sum_{i=1}^M \{C_i^I I_i(n) + C_i^J [J_i(n)]^+ + C_i^B [-J_i(n)]^+ + B_i H(J_i(n) < 0)\} \quad (12)$$

で与えられる。ここで $H(e)$ は、事象 e が起これば値 1 を、起こらなければ値 0 をとる定義関数

である。

g を 1 期当たりの平均費用、 $h(s_n)$ を相対費用とおけば、次の最適性方程式が成り立つ。

$$g + h(s_n) = \min_{a \in K(s_n)} \left\{ r(s_n, a) + \sum_{s_{n+1} \in S} p(s_n, s_{n+1}, a) h(s_{n+1}) \right\}, \quad s_n \in S \quad (13)$$

最適政策は、各 s_n で(13)式右辺を最小化する決定として定められる。ここで、相対費用 $h(s)$ は適当に定められた状態 s_r で $h(s_r) = 0$ である[27, 28]。

4. 修正政策反復法と SBMPIM

最適性方程式(13)を解くアルゴリズムが政策反復法 (PIM) [27]であり、修正政策反復法 (MPIM)である。特に、MPIM は PIM の値決定ルーチンを有限回の反復で置き換えた手法であり、比較的規模の大きな問題に対しても有効である[29,30]。以下、説明の便宜上用いてきた前節の s_n , s_{n+1} を s , s' で置きかえる。

[MPIM]

ステップ 1: $h^0(s_r) = 0$ をみたま初期ベクトル h^0 , 非負整数 m , 初期政策 f^0 , 正数 ε を定め, $k=0$ とおく。

ステップ 2: (政策改良ルーチン) 各 $s \in S$ に対して、

$$g^{k+1}(s) = \min_{a \in K(s)} \left\{ r(s, a) + \sum_{s' \in S} p(s, s', a) h^k(s') - h^k(s) \right\} \quad (14)$$

を計算し、 $f^k(s)$ が $g^{k+1}(s)$ を与えれば、 $f^{k+1}(s) = f^k(s)$ とおき、さもなければ、 $g^{k+1}(s)$ を与える任意の決定を $f^{k+1}(s)$ ととる。

ステップ 3: (値近似ルーチン)

$$w^0(s) = h^k(s) + g^{k+1}(s), \quad s \in S \quad (15)$$

とおき、 $l=0, 1, \dots, m-1$ に対して順次、

$$w^{l+1}(s) = r(s, f^{k+1}(s)) + \sum_{s' \in S} p(s, s', f^{k+1}(s)) w^l(s'), \quad s \in S \quad (16)$$

を計算し、

$$h^{k+1}(s) = w^m(s) - w^m(s_r), \quad s \in S \quad (17)$$

とおく。すべての s に対して、 $|h^{k+1}(s) - h^k(s)| < \varepsilon$ であれば終了。さもなければ、 $k=k+1$ として、ステップ 2 へ。

生産ラインの最適制御問題を MPIM で解くことを考える。簡単のため、 $i=1, \dots, M$ にたいし

て $L_i=1$, $I_{\max,i}=I_{\max}$, $J_{\max,i}=J_{\max}$ とおき, 品切れの最大値を B_{\max} とおけば, 状態空間 S の要素数は $(I_{\max}+1)^M (J_{\max}+I_{\max}+1)^{M-1} (J_{\max}+B_{\max}+1)$ である. 例えば, $I_{\max}=J_{\max}=B_{\max}=9$, のとき 190^M である. したがって, 状態空間 S の全ての状態にたいして値近似ルーチンを実行することは実際的ではなく, シミュレーションを用いることが考えられる. すなわち, 実際によく生起する初期状態 s_0 から出発し, システムの状態変化と費用をシミュレートし, 訪問した状態 s にたいしてだけ相対費用 $h(s)$ を推定する. この NDP アルゴリズムを SBMPIM (Simulation-Based Modified Policy Iteration Method) と呼ぶことにする.

[SBMPIM]

ステップ1: 初期状態 s_0 と望ましい状態 s^* を定め, シミュレーション回数 m および λ ($0 \leq \lambda \leq 1$) を定めて, 訪問した状態の集合 $S_v = S_T = \emptyset$ (空集合), 累積費用 $TC=0$, $s=s_0$, $k=l=1$ とおく.

ステップ2: $s \notin S_v$ ならば, $S_v = S_v \cup \{s\}$, $S_T = S_T \cup \{s\}$, s の訪問回数 $v(s)=1$ とおき, $f(s)$ を状態 s^* へ向かう実行可能な決定と定め, $u(s)=r(s, f(s))$ とおく. $s \in S_v$ ならば, $s \in S_T$ のとき, $S_T = S_T \cup \{s\}$, $v(s)=1$, $u(s)=r(s, f(s))$ とおき, $s \in S_T$ ならば,

$$\begin{aligned} v(s) &= v(s) + 1, \\ u(s) &= u(s) + r(s, f(s)) \end{aligned}$$

と更新する. 状態 s で決定 $f(s)$ をとったときの状態推移をシミュレーションし, 次期の状態 s' を定める.

$$\begin{aligned} TC &= TC + r(s, f(s)) \\ s &= s' \end{aligned}$$

と更新し, $l=m$ ならばステップ3へ. さもないければ $l=l+1$ としてステップ2へ.

ステップ3: (g の推定) 平均費用 g を次式により推定する.

$$g = TC/m$$

ステップ4: ($h(s)$ の推定) S_v のなかで s_r を定め,

$h(s_r) = (1 - \lambda v(s_r)/m)(w(s_r) - g) + \lambda v(s_r)/m(u(s_r)/v(s_r) - g)$ を計算し, $s(\neq s_r) \in S_v$ にたいして

$h(s) = (1 - \lambda v(s)/m)(w(s) - g) + \lambda v(s)/m(u(s)/v(s) - g) - h(s_r)$ を計算し, $h(s_r)=0$ とおく. ただし, $k=1$ のとき

には $h(s_r) = u(s_r)/v(s_r) - g$,

$h(s) = u(s)/v(s) - g - h(s_r)$ である.

ステップ5: (政策改良ルーチン) $s \in S_v$ にたいして

$$w(s) = \min_{a \in N(s, f(s))} \left\{ r(s, a) + \sum_{s' \in S} p(s, s', a) h(s') \right\}$$

を計算し, $v(s)=1$ とおく. ここで $N(s, f(s))$ は $K(s)$ における $f(s)$ の近傍であり, $p(s, s', a) > 0$ となる $s' \in S_v$ にたいしては, $S_v = S_v \cup \{s'\}$, $v(s')=1$ とおき, $f(s')$ を s^* へ向かう実行可能な決定と定める. $w(s') = r(s', f(s'))$ とおき,

$$h(s') = h(s)$$

として, $w(s)$ を計算する. $f(s)$ が $w(s)$ を与えなければ, $w(s)$ を与える任意の決定として $f(s)$ を改良する. k が停止回数に達すれば終了. さもないければ $S_T = \emptyset$, $TC=0$, $l=1$, $k=k+1$ とおきステップ2へ.

4. SMART と SBPI アルゴリズム

MPIM と SBMPIM を最適制御問題に適用するに先立ち, 既存の NDP アルゴリズムを簡単に紹介する.

[SMART] [31]

ステップ1: 全ての $s \in S$ と $a \in K(s)$ にたいして Q-factor $Q_{new}(s, a) = Q_{old}(s, a) = 0$, 累積費用 $TC=0$, 累積時間 $T=0$, 平均費用 $g=0$, 反復回数 $k=0$ とおき, パラメータ $(\alpha_0, \alpha_r, p_0, p_r)$ を初期化する. ステップ2: 反復 k で状態 s にいれば, α_k , p_k を

$$\begin{aligned} \alpha_k &= \alpha_0(\alpha_r + k) / (k^2 + k + \alpha_r), \\ p_k &= p_0(p_r + k) / (k^2 + k + p_r) \end{aligned}$$

として定める.

ステップ3: 高い確率 $(1-p_k)$ で $Q_{new}(s, a)$ を最小にする決定 a^* を選択し, 確率 p_k で a^* を除く $K(s)$ からランダムに a を選択する.

ステップ4: 選択された決定 a でシミュレーションを行い, 状態 s' へ推移すれば, 直接費用 $r(s, s', a)$ がかかる.

ステップ5: $Q_{new}(s, a)$ を次式により更新する.

$$\begin{aligned} Q_{new}(s, a) &= (1 - \alpha_k) Q_{old}(s, a) \\ &+ \alpha_k \left\{ r(s, s', a) - g + \min_{a' \in K(s')} Q_{old}(s', a') \right\} \end{aligned}$$

ステップ6: ステップ3で決定 a^* を選択したならば, TC と g を更新する.

$$TC = TC + r(s, s', a^*)$$

$$T = T + 1$$

$$g = TC/T$$

ステップ7: $Q_{old}(s, a) = Q_{new}(s, a)$ と更新する.

ステップ8: k が停止回数に達すれば終了. さもないければ $k=k+1$, s' を s としてステップ2へ.

Gosavi[32]は SMART が必ずしも収束しないことを示し, その改良版 RELAXED-SMART を提案している. SMART と RELAXED-SMART は SMDP にたいするアルゴリズムであるが, ここでは MDP にたいするものに修正している. [RELAXED-SMART] [32]

ステップ3~5, 7, 8 は[SMART]と同じである.

ステップ1: Q-factor $Q_{new}(s, a) = Q_{old}(s, a) = 0$, $TC = 0$, $T = 0$, $g = 0$, $k = 0$ とおき, パラメータ α_0 , p_0 , β_0 を与える.

ステップ2: 反復 k で状態 s にいれば, α_k , p_k , β_k を

$$\alpha_k = \alpha_0/k, \quad p_k = p_0/k, \quad \beta_k = \beta_0/k$$

として定める.

ステップ6: ステップ3で決定 a^* を選択したならば, TC , T , g を次式で更新する.

$$TC = (1 - \beta_k)TC + \beta_k r(s, s', a^*)$$

$$T = (1 - \beta_k)T + \beta_k$$

$$g = TC/T$$

一方, He 等[33]は PIM の値決定ルーチンをシミュレーションで置きかえた SBPI (Simulation Based Policy Iteration) アルゴリズムを提案している.

[SBPI アルゴリズム] [33]

ステップ1: 初期政策 $\{f^0(s); s \in S\}$ を定め, $k = 0$ とおく.

ステップ2: (値決定ルーチン)

2-a: (g^k の推定)

i) 初期状態 s_0 からシミュレーションにより s_1, \dots, s_m を生成する.

ii) $g^k = 0$ とおき, $n = 0, \dots, m-1$ にたいして (s_n, s_{n+1}) の推移に伴う g^k を次式で更新する.

$$g^k = (1 - 1/(n+1))g^k + (1/(n+1))r(s_n, s_{n+1}, f^k(s_n))$$

2-b: ($h^k(s)$ の推定)

i) 再帰状態 s^* を上記ステップ2-a i)で訪問回数最大の状態とする.

ii) 過渡状態 s_0 から出発し, 状態 s^* へ至るトラジェクトリーをシミュレーションにより L 本生成する.

iii) L 本目のトラジェクトリ $(s_0, s_1, \dots, s_N = s^*)$, $l = 1, \dots, L$, にたいして, 推移 (s_n, s_{n+1}) に伴う $w(s_i)$, $i = 1, \dots, n$, を次式により更新する.

$$w(s_i) = w(s_i) + \gamma_i \lambda^{n-i} d_n$$

ここで, γ_i はそのトラジェクトリ中で s_i を訪問した回数の逆数であり, $0 \leq \lambda \leq 1$,

$$d_n = r(s_n, s_{n+1}, f^k(s_n)) - g^k + w(s_{n+1}) - w(s_n)$$

である.

$$\text{iv) } h^k(s) = w(s) - w(s_r), \quad s \in S$$

ステップ3: (政策改良ルーチン)

$$f^{k+1}(s) = \arg \min_{a \in K(s)} \left\{ r(s, a) + \sum_{s' \in S} p(s, s', a) h^k(s') \right\}, \quad s \in S$$

ステップ4: $f^{k+1}(s) = f^k(s)$, $s \in S$ ならば停止. 最適政策は $f^k(s)$ である. さもないければ $k=k+1$ としてステップ2へ.

5. 数値実験

SMART, RELAXED-SMART, SBPI と SBMPIM アルゴリズムを3章における多工程生産ラインの最適制御問題へ適用する. まず, 単一工程生産ラインにたいし, SBMPIM を適用し, SMART, RELAXED-SMART, SBPI に対する結果をも含めて MPIM による厳密解との比較を行う. 次に, 2工程生産ラインにたいしてこれら NDP アルゴリズムを適用し, 数値比較を行う. ついで, 単一工程および2工程生産ラインをかんぱん枚数を最適に設定した JIT 生産ラインとして運用した平均費用と数値比較を行い, 最適制御による効果を明らかにする.

パラメータを以下のように設定する.

$$M = L = 1, \quad I_{\max} = 10, \quad J_{\max} = 10, \quad B_{\max} = 10,$$

$$C = 7, \quad C_{\min} = 5, \quad C^I = 1, \quad C^J = 2, \quad C^B = 5, \quad B = 10$$

そして, 工程故障を考慮した生産能力分布

$$P(C(n) = c) = P_c, \quad C_{\min} \leq c \leq C \text{ として } P_7 = 0.6,$$

$P_6 = 0.3$, $P_5 = 0.1$ とし, 需要 $D_n, n = 1, 2, \dots$ の分布は, 変形した二項分布

$$\Pr\{D_n = D - \frac{1}{2}Q + j\} = \binom{Q}{j} \left(\frac{1}{2}\right)^Q, \quad 0 \leq j \leq Q$$

ここで, D は整数, Q は偶数 ($Q \leq 2D$) であり, 分布の平均 D , 分散は $Q/4$ である. 以下の数値例では $D = 5$, $Q = 4$ を用いる.

上記パラメータ設定のもとでの各アルゴリズムの計算時間および最終の平均費用 g を表1に示す. ここで MPIM の反復回数は 63 回であった. 計算機は DOS/V 機 (CPU: Athlon 1.1GHz, メモリ: 512MB) を用いた. なお, SBMPIM の平均費用はステップ3の g にバッチ平均法 (例えば [34], p.97) を適用して求めた 95% 信頼区間である.

表2は、MPIMによる最適政策のもとでの再帰状態における、各アルゴリズムの最終政策をまとめたものである。MPIMの列に最適政策が示されており、第1列が発注量、第2列が生産量を示している。各NDPアルゴリズムにおける第3列の○は最適政策と一致したことを示している。これらからSBMPIMが他のNDPアルゴリズムより優れていることは明らかである。

同じ単一工程をかんばん方式で運用した際の最小平均費用をシミュレーション[35]により求めた。バッチサイズ 10^4 、バッチ数30のバッチ平均法による結果は、引き取りかんばん10枚、生産指示かんばん10枚のとき、平均費用が最小となり、 31.110 ± 0.801 であった。すなわち、MPIMによる最適制御にくらべ、最適かんばん方式の平均費用は約5%増加する。また、最適かんばん方式の発注量、生産量は、おおむね最適政策と一致する。しかし、引き取りかんばん、生産指示かんばん共に10枚であるので、状態(4, 7)では発注量6、生産量3、状態(4, 8)では発注量6、生産量2になる等、最適生産量を下回る場合や、状態(4, 10)では発注量6、生産量0、状態(5, 10)では発注量5、生産量0と最適発注量を上回る場合等がある。

ついで、2工程に対してSBMPIMを適用する。すなわち、

$$\begin{aligned} M=2, L=1, (I_{\max:1}, I_{\max:2}) &= (10, 10), \\ (J_{\max:1}, J_{\max:2}) &= (10, 10), B_{\max} = 10, \\ (C_1, C_2) &= (7, 7), (A_1^I, A_2^I) = (1, 3), \\ (B_1^I, B_2^I) &= (2, 5), (C_1^B, C_2^B) = (10, 10), \\ (B_1, B_2) &= (20, 20) \end{aligned}$$

であり、需要分布は単一工程と同じ分布を用いることにする。工程1, 2の生産能力分布としては、

A: $P_7 = 1$ (故障なし), 平均生産能力=7

B: $P_7 = 0.6$, $P_6 = 0.3$, $P_5 = 0.1$, 平均生産能力=6.5

の2分布を考える。ここでBは上記単一工程の生産能力分布である。表3は、生産能力分布A, BにたいしてSBMPIMによる平均費用と計算時間および最適かんばん方式の平均費用とかんばん枚数(工程1の引き取り、生産指示かんばん枚数、工程2の引き取り、生産指示かんばん枚数)を示している。最適かんばん方式はAにたいしてSBMPIMの準最適政策による平均費用を約6%以上増加させ、Bにたいしては

少なくとも同等以上の平均費用を要する。なお、最適かんばん方式が20%以上平均費用を増加させる数値例も得られている。表4はA, Bにたいしてえられた準最適政策である最終政策の変化を示している。

6. おわりに

近年、米国を中心とした、Solelectronをはじめとする電子機器製造を請け負うEMSが日本へ進出している。これらの武器は、全世界を対象に、グローバル最適調達を実現する情報ネットワークにある。日本製造業の復権をめざすためには、これに加えて、POSとPOP情報を適切に利用し、全体最適をめざした情報システムを構築しなければならない。筆者が[36]で、「JIT生産システムは、徹底的なムダの排除によるコスト低減をめざした生産システムである。そして、「ムダ」もまた、時代の変化と共に変化し、「かんばん」も「ムダ」になる時代がくるのかもしれない。」と記して数年、「かんばん」も「ムダ」になったのかもしれない。

参考文献

- [1] 大野勝久, 「生産ラインの最適制御」, オペレーションズ・リサーチ, Vol. 47, No. 4(「JIT生産システムの数理」特集号), pp.237-242, 2002.
- [2] J. A. Buzacott and I. G. Shanthikumar, *Stochastic Models of Manufacturing Systems*, Prentice Hall, NJ, 1993.
- [3] R.S.Sutton and A.G.Barto: Reinforcement Learning, MIT Press (1998) (三上, 皆川共訳「強化学習」, 森北出版, 2000)
- [4] D.P.Bertsekas and J.N. Tsitsiklis: *Neuro-Dynamic Programming*, Athena Scientific (1996)
- [5] R. V. Roy: "Neuro-dynamic programming: overview and recent trends," pp.431-459, in E. A. Feinberg and A. Schwartz ed. *Handbook of Markov Decision Processes*, Kluwer Academic Publishers (2002)
- [6] A. J. Clark and H. Scarf, "Optimal policies for multi-echelon inventory problem," *Management Science*, Vol. 6, pp.475-490, 1960.
- [7] L. A. Johnson and D. C. Montgomery, *Operations Research in Production Planning, Scheduling and Inventory Control*, John Wiley & Sons, NY, 1974.
- [8] 小島貢利, 中島健一, 「かんばん方式の数理」, 特集号 pp.225-230.
- [9] M. L. Spearman, D. L. Woodruff and W. J.

- Hopp, "CONWIP: A pull alternative to Kanban," *International Journal of Production Research*, Vol.28, No. 5, pp.879-894, 1990.
- [10] 大野勝久, 「待ち行列システムのスケジューリング」, システム/制御/情報, Vol. 41, No. 4, pp.117-122, 1997.
- [11] K. Ohno and K. Ichiki, Computing optimal policies for controlled tandem queueing systems, *Operations Research*, Vol. 35, No. 1, pp.121-126, 1987.
- [12] R. V. Evans, "Inventory control of a multiproduct system with a limited production resource," *Naval Research Logistics Quarterly*, Vol. 14, pp. 173-184, 1967.
- [13] G. A. Decroix and A. Arreola-Risa, "Optimal production and inventory policy for multiple products under resource constraints," *Management Science*, Vol. 44, No. 7, pp. 950-961, 1998.
- [14] R. Rubio and L. M. Wein, "Base stock levels using product-form queueing networks," *Management Science*, Vol. 42, No. 2, pp. 259-268, 1996.
- [15] P. Glasserman and S. Tayur, "The stability of capacitated, multi-echelon production-inventory system under a base-stock policy," *Operations Research*, Vol. 42, No. 5, pp. 913-925, 1994.
- [16] F. Chen, "Echelon reorder points, installation reorder points, and the value of centralized demand information," *Management Science*, Vol. 44, No. 12, pp. S221-S234, 1998.
- [17] M. L. Spearman and M. A. Zazanis, "Push and pull production systems: issues and comparisons," *Operations Research*, Vol. 40, No. 3, pp.521-532, 1992.
- [18] J. A. Muckstadt and S. R. Tayur, "A comparison of alternative kanban control mechanisms I," *IIE Transaction*, Vol. 27, No. 2, pp. 140-150, 1995.
- [19] J. A. Muckstadt and S. R. Tayur, "A comparison of alternative kanban control mechanisms II," *IIE Transaction*, Vol. 27, No. 2, pp. 151-161, 1995.
- [20] K. K. Yang, "Managing a flow line with single-kanban, dual-kanban or CONWIP," *Production and Operations Management*, Vol. 9, No. 4, pp. 349-366, 2000.
- [21] R. Akila and P. R. Kumar, "Optimal control of production rate in a failure prone manufacturing system," *IEEE Transactions on Automatic Control*, Vol. AC-31, No. 2, pp. 116-126, 1986.
- [22] T. Bielecki and P. R. Kumar, "Optimality of zero-inventory policies for unreliable manufacturing systems," *Operations Research*, Vol. 36, No. 4, pp.532-541, 1988.
- [23] G. V. Ryzin and S. B. Gershwin, "Production control for a tandem two-machine system," *IIE Transactions*, Vol. 25, No. 5, pp. 5-20, 1993.
- [24] M. H. Veatch and L. M. Wein, "Optimal control of a two-station tandem production/inventory system," *Operations Research*, Vol. 42, No. 2, pp. 337-350, 1994.
- [25] K. Ohno and K. Nakashima, "Optimality of a Just-in-Time production system," *Proceedings of APOPS'94*, pp.390-398, World Scientific, 1995.
- [26] K. Ohno, K. Nakashima and M. Kojima, "Sub-optimality of a JIT production system with stochastic demand," *Japan/USA Symposium on Flexible Automation*, Vol.2, pp.1253-1256, ASME, 1996.
- [27] R. A. ハワード, 「ダイナミック プログラミングとマルコフ過程」 培風館, 1971.
- [28] M. L. Puterman, *Markov Decision Process*, John Wiley & Sons, 1994.
- [29] K. Ohno, "Modified policy iteration algorithm with nonoptimality tests for undiscounted Markov decision process," Working Paper, Dept. of Information System and Management Science, Konan University, Japan, 1985.
- [30] 大野勝久: "マルコフ決定過程", システムと制御, Vol. 29, No. 6, pp.333-341 (1985)
- [31] T. K. Das, A. Gosavi, S. Mahadevan and Nich. Marchallick: "Solving semi-Markov decision problem using average reward reinforcement learning", *Management Science*, Vol. 45, No.4, pp.560-574 (1999)
- [32] A. Gosavi: Doctor Thesis, <http://faculty.uscolo.edu/gosavi/thesis.html> (1999)
- [33] Y. He, M. C. Fu and S. I. Marcus: "A Simulation-based policy iteration algorithm for average cost unichain Markov decision processes", M. Laguna and J. L. G. Velarde ed, *Computing Tools for Modeling, Optimization and Simulation*, Kluwer Academic, pp.161-182 (2000)
- [34] 大野勝久, 田村隆善, 森健一, 中島健一: 「生産管理システム」, 朝倉書店 (2002)
- [35] 石塚陽, 山下英明: 「サンプルパス最適化の確率的離散事象システムへの適用」, オペレーションズ・リサーチ, Vol. 46, No. 4, pp.195-201 (2001)
- [36] 大野勝久, 「JIT 生産システムにおけるスケジューリング」, 第 40 回システム制御情報学会研究発表講演会講演論文集, pp.39-44, 1996.

表1 計算時間 (単位は秒)

解法	計算時間 (秒)	平均費用 (g)
MPIM	2.24	28.806
SBMPIM	12.58	28.576±1.756
SMART	31.55	59.975
RELAXED-SMART	41.48	63.763
SBPI	0.97	63.920

表2 各 NDP アルゴリズムによる最終政策の比較

部品 在庫量	製品 在庫量	MPIM		SBMPIM			SMART			RELAXED- SMART			SBPI		
4	0	6	4	6	4	○	0	0		0	0		0	0	
4	1	6	4	6	4	○	0	0		0	0		0	0	
4	2	6	4	6	4	○	0	0		0	0		0	0	
4	3	6	4	6	4	○	0	0		0	0		0	0	
4	4	6	4	6	4	○	0	0		0	0		0	0	
4	5	6	4	6	4	○	0	0		0	0		0	0	
4	6	6	4	6	4	○	0	0		0	0		0	0	
4	7	6	4	6	4	○	0	0		0	0		0	0	
4	8	6	4	6	4	○	0	0		0	0		0	0	
4	9	6	4	6	4	○	0	0		0	0		0	0	
4	10	5	3	5	3	○	0	0		0	0		0	0	
5	0	5	5	5	5	○	0	0		0	0		0	0	
5	1	5	5	5	5	○	0	0		0	0		0	0	
5	2	5	5	5	5	○	0	0		0	0		0	0	
5	3	5	5	5	5	○	0	0		0	0		0	0	
5	4	5	5	5	5	○	0	0		0	0		0	0	
5	5	5	5	5	5	○	0	0		0	0		0	0	
5	6	5	5	5	5	○	0	0		0	0		0	0	
5	7	5	5	4	5		0	0		0	0		0	0	
5	8	5	5	4	5		0	0		0	0		0	0	
5	9	5	4	5	4	○	0	0		0	0		0	0	
5	10	4	3	4	3	○	0	0		0	0		0	0	

○は MPIM と政策が一致したことを示している。

表3 生産能力分布の変化に伴う準最適制御 (SBMPIM) と最適かんばん方式の比較

生産能力分布	SBMPIM	SBMPIM 計算時間(分)	最適かんばん方式 最適かんばん枚数		かんばん方式/SBMPIM
A	68.196±2.427	20.033	76.550±1.789	10-6-10-10	≥105.859
B	71.874±2.691	156.379	76.249±1.686	10-5-10-10	≥99.997

表4 生産能力分布による準最適政策の変化

状態				A				B			
部品 1	製品 1	部品 2	製品 2	発注 1	生産 1	発注 2	生産 2	発注 1	生産 1	発注 2	生産 2
4	1	5	2	6	4	5	5	6	4	5	5
4	1	5	3	6	4	5	5	6	4	5	5
4	1	5	4	6	4	5	5	6	4	5	5
4	1	5	5	6	4	5	5	6	4	5	5
4	1	5	6	6	4	5	5	6	4	5	5
5	0	5	0	5	5	5	5	5	5	5	5
5	0	5	1	5	5	5	5	5	5	5	5
5	0	5	2	5	5	5	5	5	5	5	5
5	0	5	3	5	5	5	5	5	5	5	5
5	0	5	4	5	5	5	5	5	5	5	5
5	0	5	5	5	5	5	5	5	5	5	5
5	0	5	6	5	5	5	4	5	5	5	5
5	1	5	0	5	5	5	5	5	5	5	5
5	1	5	1	5	5	5	5	5	5	5	5
5	1	5	2	5	5	5	5	5	5	5	5
5	1	5	3	5	5	5	5	5	5	5	5
5	1	5	4	5	5	5	5	5	5	5	5
5	1	5	5	5	4	5	5	5	5	5	5
5	1	5	6	5	4	5	5	5	5	5	5
6	0	5	2	4	6	5	5	4	6	5	5
6	0	5	3	4	6	5	5	4	6	5	5
6	0	5	4	4	6	5	5	4	6	5	5
6	0	5	5	4	6	5	5	4	6	5	5
6	0	5	6	4	6	5	5	4	6	5	5